Quasi-Hamiltonian Equations of Motion for Internal Coordinate Molecular Dynamics of Polymers

ALEXEY K. MAZUR[1,2]

[1] *Laboratoire de Biochimie Théorique, CNRS UPR9080, Institute de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, Paris 75005, France*

[2] *Pacific Institute of Bioorganic Chemistry, Russian Academy of Sciences Vladivostok 690022, Russia*

Received 6 September 1996; accepted 22 February 1997

ABSTRACT: Conventional molecular dynamics simulations of macromolecules require long computational times because the most interesting motions are very slow compared to the fast oscillations of bond lengths and bond angles that limit the integration time step. Simulation of dynamics in the space of internal coordinates, that is, with bond lengths, bond angles, and torsions as independent variables, gives a theoretical possibility of eliminating all uninteresting fast degrees of freedom from the system. This article presents a new method for internal coordinate molecular dynamics simulations of macromolecules. Equations of motion are derived that are applicable to branched chain molecules with any number of internal degrees of freedom. Equations use the canonical variables and they are much simpler than existing analogs. In the numerical tests the internal coordinate dynamics are compared with the traditional Cartesian coordinate molecular dynamics in simulations of a 56 residue globular protein. For the first time it was possible to compare the two alternative methods on identical molecular models in conventional quality tests. It is shown that the traditional and internal coordinate dynamics require the same time step size for the same accuracy and that in the standard geometry approximation of amino acids, that is, with fixed bond lengths, bond angles, and rigid aromatic groups, the characteristic step size is 4 fs, which is 2 times higher than with fixed bond lengths only. The step size can be increased up to 11 fs when rotation of hydrogen atoms is suppressed. *J Comput Chem* **18:** 1354–1364, 1997

Keywords: molecular dynamics of polymers; internal coordinate molecular dynamics; constraint dynamics; molecular dynamics in torsion angle space; implicit leapfrog integrator

 CCC 0192-8651 / 97 / 111354-11

Introduction

Computer simulations of molecular dynamics (MD) have become an invaluable tool applied to numerous problems in chemical physics[1] and biophysics.[2] Normally molecular movements are simulated by the method of point particles,[3] that is, the motion of each atom is computed by Newton's equations while molecular structures are maintained by harmonic potentials that keep bond lengths and bond angles close to their standard values. The maximum time step with which Newton's equations can be numerically integrated is limited to rather small values (1 fs for systems containing explicit hydrogen atoms) in order to reproduce bond-length oscillations accurately. This time step is annoyingly small for simulations of macromolecules, notably biopolymers. The unique molecular properties of biopolymers result from specific concerted atom movements (conformational transitions) that can occur over a very broad time range of picoseconds to seconds. Because of the necessarily small time step, interesting motions mostly remain beyond the reach of even the fastest computers. This presents the central problem in the methodology in the field.

Twenty years ago Ryckaert et al.[4] proposed an approach to this problem that is now called constraint dynamics. They found a simple and efficient way to impose holonomic constraints upon atom-atom distances in a system of point particles governed by Newton's equations. Their well-known algorithm, SHAKE, and a few other techniques based upon the same idea[5a,b] are now often used to fix bond lengths, increasing the time step limit to 2 fs for systems with explicit hydrogens. Constraint dynamics has been extensively reviewed.[1,6,7] These simple algorithms can be applied to any molecules regardless of their size, chemical structure, etc. However, although it seemed initially that the method could fix not only bond lengths but also, by triangulation, could constrain bond angles, dihedral angles, rigid planar groups, etc.,[4] this was found to be true only for very small molecules. In large complex polymers like proteins even bond angles cannot be fixed in this way.[8]

The intrinsic limitation of constraint dynamics can be qualitatively understood from the underlying physical model. Imposing a distance constraint implies that a reaction force is introduced that is applied along a line joining two atoms. Reactions must be calculated at each time step to balance all other forces in the system, which for large molecules presents a difficult problem because they are all coupled and form a system of algebraic equations solved by iterations. With only bond lengths to hydrogens, fixed reactions are coupled in small groups and iterations converge rapidly. With all bond lengths fixed, reactions are coupled globally and looping becomes possible. Generally in this case the convergence is not guaranteed,[9] but in practice it remains acceptable. When bond angles are triangulated coupling is much stronger and convergence becomes too slow. Efforts to overcome these limitations continue[10]; in particular, it was shown that an exact noniterative calculation of reactions is possible in principle,[11] but, so far, this is practical only for water molecules[12] and for linear unbranched chains in which case the problem can be reduced to inversion of a banded matrix.[5a,c]

Despite these difficulties it remains desirable to model chain molecules within the so-called standard geometry approximation,[13] that is, with standard fixed conformations of individual chemical groups and with rotations around single bonds as the only motion considered. Rotatable torsion angles completely define molecular conformations and present a convenient minimum set of generalized coordinates that are directly used in static conformational analysis and Monte Carlo simulations of polymers. In principle, MD simulations in the space of generalized, rather than Cartesian, coordinates are also possible by using one of the appropriate general formalisms of classical mechanics. In this way all limitations imposed by the fast vibrations of bond lengths, bond angles, etc., would be eliminated automatically. This approach was first applied to liquid butane with 1 internal degree of freedom[14] and soon after Pear and Weiner pioneered simulations of linear chains of up to 15 bonds based upon the Newton–Euler equations of rigid body dynamics.[15] Interest in this line of research decreased because of its complexity and because it initially appeared that constraint dynamics could solve all problems, but the difficulties encountered have led to a revival of the work.

To date several research groups have reported attempts to find a suitable technique for MD simulations in generalized coordinates.[15–20] Both methods of rigid body dynamics[15,16,19,20] and the Lagrange–Hamilton formalism[17,18] were employed. These methods often have little in com-

mon in their analytical formulations, but they all may be reasonably referred to as internal coordinate molecular dynamics (ICMD). The full set of internal coordinates includes not only torsions but also bond lengths and bond angles, and in some approaches these variables may also be free, if desired. The term ICMD emphasizes the main distinction of these approaches from conventional MD: they all consider molecular motion in the space of generalized internal coordinates rather than in the usual Cartesian coordinate space.

None of the ICMD techniques has yet been sufficiently well developed to reach its main goal, that is, to compute long-duration macromolecular trajectories with acceptable accuracy but at a lower cost than the classical MD with bond-length constraints. Compared with the pure Newtonian dynamics, there are two obvious problems with ICMD. The first is the derivation of equations of motion for large molecules. This task seemed too complicated at the beginning[4]; but several acceptable solutions have been proposed, which represent the main achievement of the early reports on the subject.[15–18] The second problem is the cost of additional calculations to be performed in order to obtain generalized accelerations. This must be low enough to be compensated by an increase in the step size. Among these calculations the necessity to invert the mass matrix of the system is an evident obstacle. This matrix is diagonal for the Newton's equations, but for internal coordinates it is a full positive definite matrix; therefore, a direct inversion scales as $O(n^3)$ with the number of degrees of freedom and this quickly becomes impractical when n exceeds ~ 100. Fortunately, a solution of this problem was found some time ago in robot mechanics,[21–23] namely, there are several recursive techniques that allow one to compute exact generalized accelerations for $O(n)$ calculations if the system can be treated as a tree of articulated rigid bodies. Two such algorithms have been employed recently[19,20] for ICMD and it was shown that, as in Newtonian dynamics, the cost of the time step can be made close to that of the evaluation of forces.[19b]

In spite of this progress the capabilities of ICMD still are not obvious because it is unclear how much the time step can be increased in practically interesting cases. The uncertainty is caused by two different reasons. First, biopolymers with only torsion degrees of freedom are very complex, heterogeneous, and essentially unharmonic systems in which it is usually impossible to distinguish the fastest motions *a priori* by intuition. For instance, in simulations of small peptides time steps can be increased up to 15-20 fs[17d,19b,24] and one might expect that a considerable increase of the step size should be possible for larger molecules as well. Surprisingly, the first simulations of torsion dynamics of proteins reported recently[19b,20] encountered a step size barrier of only 2 fs, the same as in the constraint dynamics with fixed bond lengths. This unusual effect of molecular size still has no clear explanation. Most of the uncertainty, however, results from unresolved numerical problems involved in integration of ICMD equations of motion. Their analytical form does not give a possibility of employing common symplectic MD integrators and that is why various general-purpose predictor-corrector schemes were usually employed. It is well-known, however, that in MD applications such algorithms tend to lose stability with lower step sizes. In addition, it was shown that they poorly conserve momenta when applied to equations of motion in generalized coordinates.[17d]

This study presents an attempt to overcome the numerical difficulties in ICMD reported in ref. 17d. These difficulties originate from the general form of the equations of motion that depends only upon the choice of coordinates, and so they could not be overcome by simple remedies. That is why an essentially new analytical formulation of ICMD has been developed with new equations that are essentially Hamiltonian but with a slightly different form. Surprisingly, they turned out to be much simpler than in other analytical ICMD formulations. Numerical integration in the space of conjugate Hamiltonian variables eliminates all earlier problems associated with poor conservation of momenta[17d] and results in trajectories as stable as those obtained with Newtonian MD.

This article principally presents an analytical background but includes representative tests in order to obtain reliable estimates of the step size in a few important simulation modes and to make sure that there are no hidden numerical problems that can undermine the method. It turns out that for protein torsion dynamics with all explicit hydrogens the optimal time step is 4 fs, 2 times larger than with only bond lengths fixed and the same as that for rigid water simulations.[26] This value is probably determined by fast collisions and rotations of hydrogen atoms, because without explicit hydrogens or with weighted inertia of hydrogen-only rigid bodies, the possible time step is beyond 10 fs. The computer cost of a time step for the torsion dynamics is similar to that of usual con-

straint dynamics with fixed bond lengths. Therefore, it is a method of choice for long simulations in which the accuracy of the standard geometry approximation is acceptable. It should also be stressed that ICMD differs qualitatively from the constraint dynamics in its efficiency with respect to the number of constraints: additional constraints in ICMD result in an increased speed for calculations.

Derivation of Equations

Let $\mathbf{r}_\alpha$ denote the position vector of atom α. The conformation of the molecule is specified by the set of N such vectors $\{\mathbf{r}\alpha\}$. Kinetic energy

$$K = \sum_{\alpha=1}^{N} \frac{m_\alpha \dot{\mathbf{r}}_\alpha}{2}, \tag{1}$$

where the overdot denotes the time derivative; the potential energy is

$$U = U(\{\mathbf{r}_\alpha\}). \tag{2}$$

and the Lagrangian is

$$L = K - U. \tag{3}$$

Assuming that the set of generalized coordinates is defined $\{\theta_i\}$ $i = 1, \ldots, n$, so that

$$\mathbf{r}_\alpha = \mathbf{r}_\alpha(\{\theta_i\}), \tag{4}$$

the Lagrangian equations of motion are

$$\frac{d}{dt}\frac{\partial K}{\partial \dot{\theta}_i} - \frac{\partial K}{\partial \theta_i} = -\frac{\partial U}{\partial \theta_i}. \tag{5}$$

Now consider how molecular conformation is specified by bond lengths, bond angles, and dihedrals. If in Figure 1 the Cartesian coordinates are given for atoms α, β, and γ, $\mathbf{r}_\delta$ is computed from $\mathbf{r}_\alpha$, $\mathbf{r}_\beta$, and $\mathbf{r}_\gamma$ given the values of dihedral ω, planar angle φ, and bond length l

$$\mathbf{r}_\delta = \mathbf{r}_\delta(\mathbf{r}_\alpha, \mathbf{r}_\beta, \mathbf{r}_\gamma, \omega, \varphi, l). \tag{6}$$

Similarly, $\mathbf{r}_\varepsilon$ is computed from $\mathbf{r}_\beta, \mathbf{r}_\gamma$, and $\mathbf{r}_\delta$, and so on. This procedure needs global coordinates of the first three atoms for initialization, but for the moment we skip this point to consider it later. The above process is not just an algorithm for computing $\{\mathbf{r}_\alpha\}$ from $\{\theta_i\}$; it is also a definition of a specific ordering of atoms and internal coordinates in the molecule. This ordering is indispensable for transformation of $\{\theta_i\}$ into $\{\mathbf{r}_\alpha\}$; it is implicit in eq. (4) but is revealed when derivatives $\partial \mathbf{r}_\alpha / \partial \theta_i$ need to be computed. For example, when dihedral ω in Figure 1 is varied, then by definition $\mathbf{r}_\alpha$, $\mathbf{r}_\beta$, and $\mathbf{r}_\gamma$ are not affected while $\mathbf{r}_\delta$ moves, as well as the following atoms. The infinitesimal displacements of atoms in response to variation of internal coordinates correspond to the motion of a tree: the very first three atoms form the base, internal coordinates close to the base move almost the whole molecule, and variables in a specific branch move only the upper part of that branch. Of course, there is no distinction between the base and the tip of the tree in the real motion: when the base gets its 6 degrees of freedom it becomes equivalent to any of the branches.

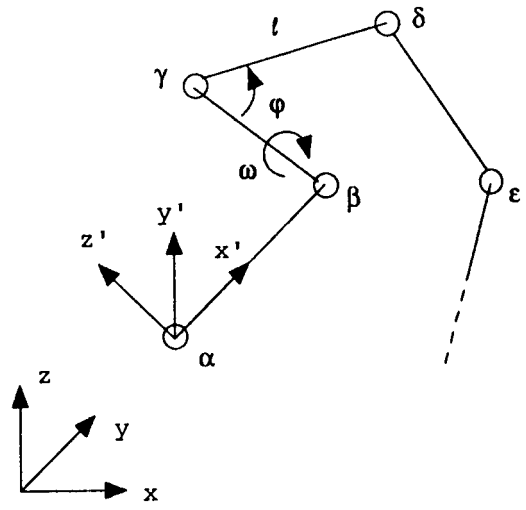

FIGURE 1. Illustration of the definition of external and internal coordinates in the molecule: α, β, γ, δ, and ε denote atom numbers; ω, φ, and l are internal coordinates used to position atom δ; xyz and $x'y'z'$ denote the global and local coordinate frames, respectively. Detailed explanations are given in the text.

Let D_i denote the set of atoms in the molecule that move when internal coordinate θ_i is varied. Then we have

$$\frac{\partial K}{\partial \theta_i} = \sum_{\alpha \in D_i} m_\alpha \dot{\mathbf{r}}_\alpha \frac{\partial \dot{\mathbf{r}}_\alpha}{\partial \theta_i}, \tag{7}$$

$$\frac{\partial K}{\partial \dot{\theta}_i} = \sum_{\alpha \in D_i} m_\alpha \dot{\mathbf{r}}_\alpha \frac{\partial \dot{\mathbf{r}}_\alpha}{\partial \dot{\theta}_i} = \sum_{\alpha \in D_i} m_\alpha \dot{\mathbf{r}}_\alpha \frac{\partial \mathbf{r}_a}{\partial \theta_i}, \tag{8}$$

and eq. (5) gives

$$\sum_{\alpha \in D_i} m_\alpha \ddot{\mathbf{r}}_\alpha \frac{\partial \mathbf{r}_\alpha}{\partial \theta_i} = -\frac{\partial U}{\partial \theta_i}. \tag{9}$$

To make the following step it is necessary to assume that there are no loops in the molecule. Internal coordinate description of flexible loops is a separate problem that needs special treatment, but without loops the whole set D_i moves as a single rigid body if only θ_i is varied. If θ_i is a bond length D_i is translated along the bond, so that for any atom

$$\delta \mathbf{r}_a = \mathbf{e}_i \, \delta\theta_i, \tag{10}$$

where $\mathbf{e}_i$ is the unit vector of the bond. If on the other hand θ_i is an angle or a dihedral, rotation of D_i occurs.

$$\delta \mathbf{r}_\alpha = \mathbf{e}_i \times (\mathbf{r}_\alpha - \mathbf{r}_i)\delta\theta_i, \tag{11}$$

where the multiplication sign denotes a vector product and the unit vector $\mathbf{e}_i$ and the position vector $\mathbf{r}_i$ specify the axis of rotation.

For a translational variable eq. (9) gives

$$\sum_{\alpha \in D_i} m_\alpha \ddot{\mathbf{r}}_\alpha \frac{\partial \mathbf{r}_\alpha}{\partial \theta_i} = \mathbf{e}_i \sum_{\alpha \in D_i} m_\alpha \ddot{\mathbf{r}}_\alpha = \mathbf{e}_i \dot{\mathbf{P}}_i = -\frac{\partial U}{\partial \theta_i} \tag{12}$$

where $\mathbf{P}_i$ is the total momentum of the set of atoms D_i. It is straightforward to show that, in the case of a translational variable, $\partial U/\partial\theta_i$ equals the sum of the forces applied to atoms of set D_i projected upon vector $\mathbf{e}_i$, and so eq. (12) is nothing but the corresponding projection of the Newton's equation. For a rotational variable we have

$$\sum_{\alpha \in D_i} m_\alpha \ddot{\mathbf{r}}_\alpha \frac{\partial \mathbf{r}_\alpha}{\partial \theta_i} = \mathbf{e}_i \sum_{\alpha \in D_i} (\mathbf{r}_\alpha - \mathbf{r}_i) \times (m_\alpha \ddot{\mathbf{r}}_\alpha) = \mathbf{e}_i \dot{\mathbf{M}}_i = -\frac{\partial U}{\partial \theta_i} \tag{13}$$

where $\dot{\mathbf{M}}_i$ is the total angular momentum of atoms of set D_i around the fixed point in space given by the current value of vector $\mathbf{r}_i$. Again it is easy to show that, for a rotational variable, $\partial U/\partial\theta_i$ equals the projection upon $\mathbf{e}_i$ of the sum of the torques around point $\mathbf{r}_i$ applied to atoms of set D_i, and therefore eq. (13) is a projection of the Newton's equations for the torques.

To shorten the notation let us introduce indicator s_i such that

$$s_i = \begin{cases} 1 & \text{if } \theta_i \text{ is rotational,} \\ 0 & \text{if } \theta_i \text{ is translational,} \end{cases} \tag{14}$$

and unite eqs. (12) and (13) as

$$s_i \mathbf{e}_i \dot{\mathbf{M}}_i + (1 - s_i)\mathbf{e}_i \dot{\mathbf{P}}_i = -\frac{\partial U}{\partial \theta_i}. \tag{15}$$

Now we need to transform eq. (15) so that a full time derivative appears on the left. We have

$$\mathbf{e}_i \dot{\mathbf{P}}_i = \frac{d}{dt}(\mathbf{e}_i \mathbf{P}_i) - \dot{\mathbf{e}}_i \mathbf{P}_i \tag{16}$$

and

$$\mathbf{e}_i \dot{\mathbf{M}}_i = \frac{d}{dt}(\mathbf{e}_i \mathbf{Q}_i) - \dot{\mathbf{e}}_i \mathbf{Q}_i + \mathbf{P}_i(\mathbf{e}_i \times \dot{\mathbf{r}}_i), \tag{17}$$

where $\dot{\mathbf{Q}}_i$ is the total angular momentum of set D_i around a moving point given by vector $\mathbf{r}_i$. Although $\mathbf{M}_i$ and $\mathbf{Q}_i$ refer to the same physical quantity, distinction between them must be made because these are two different time functions. $\mathbf{M}_i$ is angular momentum around a fixed node; it equals $\mathbf{Q}_i$ at the given moment of time, but it has a different time derivative,

$$\dot{\mathbf{M}}_i = \frac{\partial}{\partial t}\mathbf{Q}_i \bigg|_{\mathbf{r}_i = \text{const}} \tag{18}$$

and, accordingly, the two values diverge immediately afterward. The Newton's equation, eq. (13), is valid for $\mathbf{M}_i$ but not for $\mathbf{Q}_i$.

Substitution of eqs. (16) and (17) into eq. (15) gives

$$\frac{d}{dt}[s_i \mathbf{e}_i \mathbf{Q}_i + (1 - s_i)\mathbf{e}_i \mathbf{P}_i] = -\frac{\partial U}{\partial \theta_i} + s_i \dot{\mathbf{e}}_i \mathbf{Q}_i - s_i \mathbf{P}_i(\mathbf{e}_i \times \dot{\mathbf{r}}_i) + (1 - s_i)\dot{\mathbf{e}}_i \mathbf{P}_i. \tag{19}$$

Now consider more carefully the term in the brackets on the left in eq. (19). By a simple transformation we get

$$s_i \mathbf{e}_i \mathbf{Q}_i + (1 - s_i)\mathbf{e}_i \mathbf{P}_i = \sum_{\alpha \in D_i} m_\alpha \dot{\mathbf{r}}_\alpha [s_i \mathbf{e}_i \times (\mathbf{r}_\alpha - \mathbf{r}_i) + (1 - s_i)\mathbf{e}_i] = \frac{\partial \mathbf{K}}{\partial \dot{\theta}_i} = \frac{\partial \mathbf{L}}{\partial \dot{\theta}_i}. \tag{20}$$

That is, this is nothing but the conjugate momentum corresponding to variable θ_i and eq. (19) is

just one of the Hamiltonian equations. We have, therefore,

$$\sum_{k=1}^{n} a_{ik}\dot{\theta}_k = s_i \mathbf{e}_i \mathbf{Q}_i + (1 - s_i)\mathbf{e}_i \mathbf{P}_i, \qquad (21)$$

where $\{a_{ik}\}$ are the coefficients of the mass matrix of the system. Equations (19) and (21) present the resultant equations of motion to be integrated by a computer. This is done in two steps. At first the right-hand side (rhs) of eq. (19) is evaluated and the time step is made for momenta. At the same time momenta propagated from the previous step are substituted into the rhs of eq. (21) and the linear system is solved by inverting the mass matrix, which is done by using the recursive algorithm by Rodriguez et al.[23, 19a] This algorithm was developed within the context of the Newton–Euler approach to rigid body dynamics, where each its step has a physical interpretation and corresponds to a special decomposition of reactions in hinges between rigid bodies. Here these interpretations are lost, but due to the same structure of matrix $\{a_{ik}\}$, it can still be applied just as a formal mathematical procedure. The generalized velocities thus obtained are used to make the time step in generalized coordinates. A specific example of an integrator for eqs. (19) and (20) employed in the numerical tests reported below is detailed in the Appendix.

Equation (19) is rather simple and presents no computational problems. All terms on the right are familiar functions of atom coordinates and atom velocities. It is clear that in an unbranched chain molecule sets D_i can be ordered so that

$$D_1 \supset D_2 \supset \cdots D_{n-1} \supset D_n. \qquad (22)$$

Therefore, all $\mathbf{P}_i$ can be computed starting from the tip of the chain and moving to the base and at the ith variable adding only the contribution from subset D_i/D_{i+1}. To compute $\mathbf{Q}_i$ angular momenta of sets, D_i are first computed with respect to the zero of global coordinates in the same way as momenta. After that $\mathbf{Q}_i$. are computed by moving the ith node of rotation to $\mathbf{r}_i$. The total force applied to atoms of set D_i is computed similarly to $\mathbf{P}_i$ and the total torque is computed similarly to $\mathbf{Q}_i$. It can be noted that the recurrent calculations of derivatives of the conformational energy used in this field for a long time[17a, 27] are nothing but the above summation of forces and torques. Due to these algorithms, computations of forces in ICMD is in fact similar to the classical MD, contrary to the common opinion.[19, 20] Generalization of the above procedure for a branched chain is straightforward. These computations present only technical difficulties and there is no need to detail them here. The computer time necessary to evaluate the nonpotential terms in eq. (19) is negligible.

It seems reasonable to call eqs. (19) and (21) quasi-Hamiltonian because of the meaning of the parameters involved. One should note that there is already some confusion in the classification of equations of motion in internal coordinates. It certainly makes little sense to base such classification upon the method used for the derivation of equations because their final form depends only upon the choice of coordinates. For example, by substituting eq. (21) into eq. (19) we can get the earlier equations for generalized accelerations[17a] that are equivalent to any other equations for accelerations of internal coordinates.[19a] Inversely, eq. (19) may be obtained by separating a full time derivative in the corresponding equations for generalized accelerations derived by any appropriate technique.

Conservation of Momenta

Let us consider now the generalized coordinates that position in space the first three atoms in the molecule. These coordinates may be called external and they affect positions of all other atoms; therefore, $\mathbf{P}_1$ and $\mathbf{Q}_1$ are the total momentum and angular momentum around a node that is yet to be specified. In the absence of external field $\partial U/\partial\theta_i = 0$; therefore, for external variables, the rhs of eq. (19) includes only inertial terms. All these terms become zero, however, if variable θ_i can be defined so that the corresponding $\mathbf{e}_i$ and $\mathbf{r}_i$ are fixed in space. For instance, if the Cartesian coordinates of the first atom are used as the first three translational coordinates, eq. (19) gives

$$\frac{d}{dt}\mathbf{P}_x = \frac{d}{dt}\mathbf{P}_y = \frac{d}{dt}\mathbf{P}_z = 0, \qquad (23)$$

where $\mathbf{P}$ is the total momentum of the molecule. Thus, the conservation of the total momentum appears to be explicitly present in equations of motion and, more importantly, in their finite-difference approximations. The same property is implicit in the Newton's equations. On the face of it this might seem rather minor, but it appears to be somehow related with the numerical stability with large step sizes. In the Newton's case, among the principal first integrals of mechanics only the

total energy is affected by the numerical approximation errors while the conservation of momenta commonly is ''ideal''; that is, it depends only upon the length of the floating point numbers in the computer. In contrast, in the numerical integration of standard ICMD equations for generalized accelerations all first integrals are similarly affected by approximation errors. The importance of this qualitative difference was recognized in an earlier study[17d] where it was possible to significantly improve the numerical stability with large step sizes by enforcing conservation of momenta. By using eqs. (19) and (20) ideal numerical conservation can be ensured with an appropriate choice of the external coordinates, and we will see below that in this way one manages to obtain the quality of numerical trajectories comparable to that of classical MD.

The above considerations dictate the following procedure for positioning the first three atoms in a molecule. Figure 1 shows the global coordinate frame xyz and the local frame $x'y'z'$ that is bound with the molecule. The origin of the local frame coincides with the first atom (atom α in Fig. 1). The second atom (β) always rests on axis $O'x'$ and the third atom (γ) always rests in plane $x'O'y'$. The second atom moves along the $O'x'$ axis if its bond lengths (bond $\alpha - \beta$ in Fig. 1) is not fixed. Similarly, the third atom moves in the $x'O'y'$ plane if bond $\beta - \gamma$ and/or bond angle $\alpha - \beta - \gamma$ are variable. With these conventions, the 6 rigid body degrees of freedom of frame $x'y'z'$ with respect to the global coordinates xyz complement the internal degrees of freedom in the molecule to the full set required.

Frame $x'y'z'$ is considered to rotate around the zero of the global coordinates and variations of the first three generalized coordinates, $\delta\theta_1$, $\delta\theta_2$, and $\delta\theta_3$, correspond to rotations around axes x, y, and z, respectively. In this case $\dot{\theta}_1$, $\dot{\theta}_2$, and $\dot{\theta}_3$ obtained from eq. (21) are the components ω_x, ω_y, and ω_z of the angular velocity of frame $x'y'z'$. They are used to integrate the kinematic equations for the quaternion that controls orientation of frame $x'y'z'$.[28] Vector $\mathbf{Q} = \mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{Q}_3$ is the total angular momentum around the global coordinate origin, therefore, eq. (19) guarantees conservation of the angular momentum. One can note, however, that because the first atom is moved by variations $\delta\theta_1$, $\delta\theta_2$, and $\delta\theta_3$, the Cartesian coordinates of the first atom can no longer be taken as independent. However, we can still select the three global translational variables so that variations $\delta\theta_4$, $\delta\theta_5$, and $\delta\theta_6$ correspond to translations of all atoms along global axes x, y, and z. In this case $\dot{\theta}_4$, $\dot{\theta}_5$, and $\dot{\theta}_6$ obtained from eq. (21) refer to v_x, v_y, and v_z components of velocity of the point that is rigidly connected with the $x'y'z'$ frame but instantaneously coincides with the zero of coordinates as the center of rotation. The velocity of the first atom is then computed from its position vector $\mathbf{r}_1$ and vectors $\boldsymbol{\nu}$ and $\boldsymbol{\omega}$. It is clear that in this case vector $\mathbf{P} = \mathbf{P}_4 = \mathbf{P}_5 = \mathbf{P}_6$ is still the total momentum of the molecule.

Numerical Tests

The formal validity of the analytical formulation of the ICMD method described in the sections above was checked in numerous tests with biopolymers of different size and chemical structure. These calculations clearly confirmed the expected superiority of this new method over our previous formulation.[17] In particular, results of the tests obtained with enforced conservation of momenta[17d] were considerably improved, especially in regard to the magnitude of the total energy drift (data not shown). The results of a few representative calculations are shown in Figure 2. Following many previous methodological MD studies, a single protein molecule in a vacuum was considered here, and the time step comparisons were made by using the conservation of the total energy to access the accuracy of the numerical trajectory. The protein molecule chosen was the immunoglobulin binding domain of streptococcal protein G[19] (PGB, file 1pgb in the Protein Database[30]), which is a 56 residue α/β protein subunit. PGB is somewhat larger than the commonly used test proteins, but it has neither S-S bridges nor proline residues so that no complications connected with the treatment of the flexible rings are involved. Two indicators of the conservation of the total energy were used: the module of the drift and the relative fluctuation, which is computed as

$$\delta = \frac{\langle (E - \langle E \rangle)^2 \rangle^{1/2}}{\langle (K - \langle K \rangle)^2 \rangle^{1/2}}, \qquad (24)$$

where $\langle E \rangle$ and $\langle K \rangle$ are the average total and kinetic energies, respectively. The value of the drift is calculated by a linear regression during 1-ps intervals along the trajectory and averaged.

In Figure 2a-d the ICMD simulations are compared with the classical MD. For the usual MD the popular package AMBER[31a] with force field

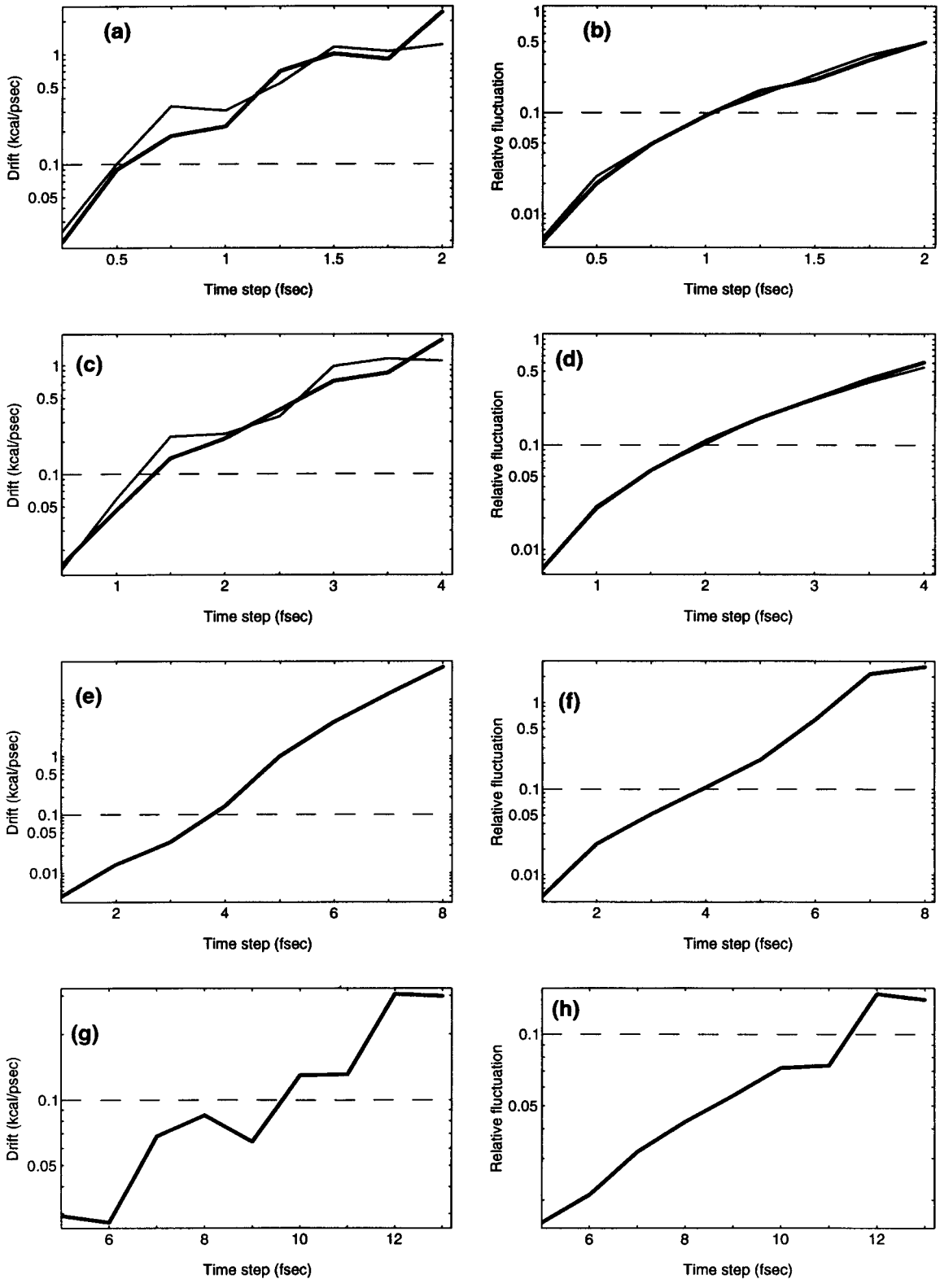

FIGURE 2. Time step dependencies of the total energy drift and the relative fluctuation for four models of PGB: (a, b) without constraints, (c, d) with fixed bond lengths to hydrogen atoms, (e, f) in the standard geometry approximation, and (g, h) in the standard geometry approximation with inertia of hydrogen-only rigid bodies increased as described in the text. Thin lines show Cartesian coordinate dynamics by AMBER. Thicker lines show internal coordinate dynamics by DY.

AMBER94[31b] was used with the distance dependent dielectric constant, $\varepsilon = r$, and no truncation of nonbonded interactions. In calculations with fixed bond lengths the SHAKE algorithm[4] was employed with a tolerance of 10^{-6}. ICMD trajectories were computed by a new program (DY) with the same force field and all internal coordinates as variables. Special care was taken to ensure the identity of the force field implementations in the two programs and the identity of the starting states. The equilibration was more or less standard and included minimization of the crystal structure followed by a 12.5-ps run starting from Maxwell distribution at 300 K with periodic temperature control and a step size of 0.5 fs. After that the step size was reduced to 0.25 fs, a short trajectory of 150 fs was calculated, and the final part was stored and used for generating initial data. The final point was used as the starting state for a test trajectory of 10 ps, and initial half-step velocities were taken at appropriate time intervals from the coordinates. These last preparations were necessary to provide smooth starts of leapfrog integrators used in AMBER and DY with different time step sizes. The same starting states were imported from AMBER to DY where internal coordinates and generalized velocities were computed from atomic coordinates and velocities. The test trajectory was repeatedly computed with gradually growing step size, first in Cartesian and then in internal coordinates.

Figure 2a, b presents the results obtained in calculations with completely free molecular models, that is, with no constraints. The agreement between the two sets of simulations is close to ideal. Note that the commonly used value of the time step (1 fs) corresponds to a relative fluctuation of $\delta = 0.1$, which can be reasonably used as a reference level of acceptable accuracy. Results of an analogous test with bond length constraints applied to hydrogen atoms are shown in Figure 2c, d. Again one can note a remarkable agreement between the two different methods. The relative fluctuation of 0.1 corresponds to the well-known time step size level of 2 fs. Similar agreement was observed in simulations with all bond lengths fixed, and it should be added that absolute values of all energy components were very close in all comparative tests (results not shown). These results demonstrate that ICMD simulations are as accurate as traditional MD and that both techniques suffer from similar time step limitations. Figure 2a-d, therefore, gives an appropriate reference to which standard geometry ICMD simulations may be compared.

The results of ICMD tests performed with the set of variables corresponding to the standard geometry approximation are shown in Figure 2e, f. Compared to Figure 2c, d, the relative fluctuation and the drift of the total energy are considerably reduced. The time step size corresponding to the relative fluctuation of 0.1 is 4 fs, which is 2 times larger than for all-atom models with fixed bond lengths. This value is close to the well-known step size limit for dynamics of rigid water molecules,[26] suggesting that the limitation is probably due to the fast rotation of hydrogen atoms in hydrogen-only rigid bodies like hydroxyl or methyl groups. One can note that the observed increase in the step size is rather moderate compared to earlier estimates obtained in simulations with smaller molecules.[17d] The apparent reason is that these fast rotations should mainly occur in protein cores, and so the test molecule must be sufficiently large and must include a large number of hydrogen-only rigid bodies for these limitations to be detected.

Figure 2g, h presents additional proof of the fact that, in the above ICMD tests, the time steps are limited by rotation of hydrogens and illustrates one simple possibility to get rid of this limitation. It is clear that rigid body rotation can be slowed down by artificially increasing its inertia. Imagine, for instance, that in ethane (CH_3–CH_3) both carbon atoms are no longer point masses but rather spheres that rotate together with the neighboring hydrogens. Rotation of hydrogens will be slowed down but the overall motion of the molecule will be perturbed only slightly because atom masses do not need to be changed. This rather general trick can be employed at all levels for balancing the time scales of different motions in any Hamiltonian system. Note, for instance, that a similar approach is applied to electron degrees of freedom in the Car–Parrinello method[32] and that the so-called weighted mass MD[33] is also based upon this idea. Results presented in Figure 2g, h were obtained with point masses of atoms in the rotation nodes of hydrogen-only rigid bodies replaced by spheres with the same mass and fixed moments of inertia equal to 15 amu $\cdot$Å^2. As a result, both the drift and the relative fluctuation are significantly reduced compared to Figure 2e, f. The time step size that corresponds to the relative fluctuation of 0.1 is about 11–12 fs.

Some comparative data on the computer cost of simulations in AMBER and DY in different modes are given in the Appendix. Overall, Table A.I shows

TABLE A.I.
Timing Comparisons.

	A[a]	B[b]	C[c]	D[d]
AMBER	1.10	1.12	1.20	—
DY	1.60	2.33	1.97	1.12

CPU time is in seconds per one step of dynamics.
[a] No constraints. Time step 1 fs.
[b] Fixed bond lengths to hydrogen atoms. Time step 2 fs.
[c] All bond lengths fixed. Time step 2 fs.
[d] Standard geometry. Time step 4 fs.

that ICMD is complementary to the traditional MD. As expected, for unconstrained simulations and for the two usual modes of bond length constraints, AMBER is somewhat faster than DY. On the other hand, in the standard geometry simulations the cost of a time step in ICMD is roughly the same as in constraint dynamics with fixed bond length to hydrogens. Thus, the observed increase in the step size is obtained essentially for no cost and in terms of computer time per picosecond, ICMD gives the best score.

Conclusions

In recent years there has been slow but steady progress in the development of techniques for ICMD. As shown here, improved ICMD methodology now makes possible simulations of torsion dynamics of biopolymers that fulfill the same strict criteria of accuracy and stability as classical MD. Overall ICMD simulations within the standard geometry approximation are at least 2 times faster than constraint dynamics.

It should be kept in mind, however, that the numerical tests presented here mainly serve to confirm the validity of the proposed equations. They do not prove that the method can be freely used in calculations of the physical properties of molecules. It is known that even constraints on bond lengths can, in certain cases, cause significant perturbations[34] and that fixing bond angles may significantly affect the conformational statistics via the so-called metric tensor.[35] Effects of each type of constraint upon the measured properties should, therefore, be thoroughly studied. These reservations, however, are not that important for many applications in structural refinement of biopolymers,[20] in conformational searches, and in simulations of protein folding.

Appendix: Implicit Leapfrog Integrator

To simplify notation let us consider the case of a single variable. Let θ and p denote the generalized coordinate and the corresponding conjugate momentum. Equation (19) may be rewritten as

$$\dot{p} = g + f \tag{A.1}$$

where $g = -\partial U/\partial\theta$ and f denotes the inertial term. Calculation of g may be denoted as $(\theta) \to g$, calculation of f as $(\theta, \dot{\theta}) \to f$, etc. The on-step values are denoted as θ_n, θ_{n+1}, etc., and the half-step values as $\theta_{n-(1/2)}, \theta_{n+(1/2)}, \ldots$, etc. Then the sequence of steps in the leapfrog integrator is expressed as

$$(\theta_n) \to g_n \tag{A.2}$$

$$p_{n+\frac{1}{2}} = p_{n-\frac{1}{2}} + (g_n + f_{n-\frac{1}{2}})h \tag{A.3}$$

$$\theta_{n+\frac{1}{2}} = \theta_{n-\frac{1}{2}} + h\dot{\theta}_{n-\frac{1}{2}} \tag{A.4}$$

$$(\theta_{n+\frac{1}{2}}, p_{n+\frac{1}{2}}) \to \dot{\theta}_{n+\frac{1}{2}} \tag{A.5}$$

$$\begin{cases} \theta_{n+\frac{1}{2}} = \theta_{n-\frac{1}{2}} + \left(\dot{\theta}_{n-\frac{1}{2}} + \dot{\theta}_{n+\frac{1}{2}}\right)\dfrac{h}{2} & \text{(A.6)} \\ \left(\theta_{n+\frac{1}{2}}, \dot{\theta}_{n+\frac{1}{2}}\right) \to f_{n+\frac{1}{2}} & \text{(A.7)} \\ p_{n+\frac{1}{2}} = p_{n-\frac{1}{2}} + g_n h + (f_{n-\frac{1}{2}} + f_{n+\frac{1}{2}})\dfrac{h}{2} & \text{(A.8)} \\ (\theta_{n+\frac{1}{2}}, p_{n+\frac{1}{2}}) \to \dot{\theta}_{n+\frac{1}{2}} & \text{(A.9)} \end{cases}$$

$$\theta_{n+1} = \theta_n + \dot{\theta}_{n+\frac{1}{2}}h, \tag{A.10}$$

where h denotes the step size.

The steps enclosed in the brace are repeated iteratively until convergence of eq. (A.6). The iterative cycle involves two stages with nontrivial calculations given by eqs. (A.7) and (A.9). The half-step coordinates are corrected in eq. (A.6); therefore, the inversion of the mass matrix is implicitly involved in eq. (A.9), which makes the algorithm relatively costly and leaves some room for possible improvements.

Some estimates of the CPU time made in the course of the numerical tests are given in Table A.I. The measurements were made on a Silicon Graphics Crimson R4000-50 workstation. DY was compiled by a standard C-compiler with the lowest level of optimization. In all calculations a relative tolerance of 10^{-6} was used as the criterion of convergence of eq. (A.6). The cost of the matrix

inversion calculations scales roughly linearly with the number of variables; therefore, it is much less expensive in the standard geometry calculations. The average number of iterations depends upon the molecular model and also upon the step size. With the optimal time steps it was minimal in the standard geometry simulations (3.9) and maximal in simulations with fixed bond lengths to hydrogens (9.5). Thus, the standard geometry approximation is the cheapest because of the reduced number of variables and faster convergence. Evaluation of forces took a similar amount of time in AMBER and in DY.

Acknowledgments

This study was initially motivated and partially based upon unpublished results obtained in collaboration with V. E. Dorofeyev.[25] I wish to thank Dr. R. Lavery for useful discussions; without his support this work could have not been finished. I would also like to thank Dr. G. Rodriguez of the Jet Propulsion Laboratory for making the set of his reprints available for me in Russia.

References

1. M. P. Allen and D. T. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, U.K., 1987.
2. C. L. Brooks III, M. Karplus, and B. M. Pettitt, *Adv. Chem. Phys.*, **71**, 175 (1988).
3. R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, McGrawHill, New York, 1981.
4. J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.*, **23**, 327 (1977).
5 (a) H. C. Andersen, *J. Comput. Phys.*, **52**, 24 (1983); (b) R. Edberg, D. J. Evans, and G. P. Morris, *J. Chem. Phys.*, **84**, 6933 (1986); (c) N. Grønbech-Jensen and S. Doniash, *J. Comput. Chem.*, **15**, 997 (1994).
6. G. Ciccotti and J. P. Ryckaert, *Comput. Phys. Rep.*, **4**, 345 (1986).
7. D. J. Tildesley, In *Computer Simulation in Chemical Physics*, M. P. Allen and D. J. Tildesley, Kluwer, London, 1993, p. 23.
8. W. F. Van Gunsteren and M. Karplus, *Macromolecules*, **15**, 1528 (1982).
9. E. Barth, K. Kuczera, B. Leimkuhler, and R. D. Skeel, *J. Comput. Chem.*, **16**, 1192 (1995).
10.. J. Durup, *J. Phys. Chem.*, **95**, 1817 (1991).
11.. J. Turner, P. Weiner, B. Robson, R. Venugopal, H. Schubelle III, and R. Singh, *J. Comput. Chem.*, **16**, 1271 (1995).
12. S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, **13**, 952 (1992).
13. N. Go and H. A. Scheraga, *J. Chem. Phys.*, **51**, 4751 (1969).
14. J. P. Ryckaert and A. Bellemans, *Chem. Phys. Lett.*, **30**, 123 (1975).
15. M. R. Pear and J. H. Weiner, *J. Chem. Phys.*, **71**, 212 (1979).
16. J. W. Perram and H. G. Petersen, *Mol. Phys.*, **65**, 861 (1988).
17. (a) A. K. Mazur and R. A. Abagyan, *J. Biomol. Struct. Dyn.*, **6**, 815 (1989); (b) R. A. Abagyan and A. K. Mazur, *J. Biomol. Struct. Dyn.*, **6**, 833 (1989); (c) A. K. Mazur, V. E. Dorofeyev and R. A. Abagyan, *J. Comput. Phys.*, **92**, 261 (1991); (d) V. E. Dorofeyev and A. K. Mazur, *J. Comput. Phys.*, **107**, 359 (1993).
18. K. D. Gibson and H. A. Scheraga, *J. Comput. Chem.*, **11**, 468 (1990).
19. (a) A. Jain, N. Vaidehi, and G. Rodriguez, *J. Comput. Phys.*, **106**, 258 (1993); (b) A. M. Mathiowetz, A. Jain, N. Karasawa, and W. A. Goddard, *Proteins: Struct. Funct. Genet.*, **20**, 227 (1994).
20. L. M. Rice and A. T. Brunger, *Proteins: Struct. Funct. Genet.*, **19**, 277 (1994).
21. A. F. Vereshchagin, *Eng. Cybernet.*, **6**, 1343 (1974).
22. R. Featherstone, *Robot Dynamics Algorithms*, Kluwer, Boston, 1987.
23. (a) G. Rodriguez, *IEEE J. Robot. Automat.*, **RA-3**, 624 (1987); (b) G. Rodriguez and K. Kreutz, *IEEE Trans. Robot. Automat.*, **8**, 65 (1992).
24. (a) V. E. Dorofeyev and A. K. Mazur, *J. Biomol. Struct. Dyn.*, **10**, 143 (1993); (b) M. G. Petukhov, A. K. Mazur and L. A. Elyakova, *Carbohydr. Res.*, **279**, 41 (1995).
25. V. E. Dorofeyev, Ph.D. thesis, Pacific Institute of Bioorganic Chemistry, Vladivostok, Russia, 1991.
26. D. Fincham, *Mol. Simul.*, **8**, 165 (1992).
27. (a) Noguti T. and N. Go, *J. Phys. Soc. Jpn.*, **52**, 3685 (1983); (b) H. Abe, W. Braun, T. Noguti, and N. Go, *Comput. Chem.*, **8**, 239 (1984).
28. (a) V. N. Branets and I. P. Shmyglevskii, *Applications of Quaternions in Rigid Body Orientation Problems*, Nauka, Moscow, 1973; (b) D. J. Evans, *Mol. Phys.*, **34**, 317 (1977).
29. T. Gallagher, P. Alexander, P. Bryan, and G. L. Gilliland, *Biochemistry*, **33**, 4721 (1994).
30. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).
31. (a) D. A. Pearlman, D. A. Case, J. C. Caldwell, G. L. Seibel, U. Ch. Singh, P. Weiner, and P. A. Kollman, *AMBER 4.0*, University of California, San Francisco, 1991; (b) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179 (1995).
32. G. Galli and M. Parrinello, In *Computer Simulations in Material Science*, M. Meyer and V. Ponitikis, Eds., Kluwer, Dordrecht, Holland, 1991, p. 283.
33. (a) R. Pomes and J. A. McCammon, *Chem. Phys. Lett.*, **166**, 425 (1990); (b) B. Mao, G. M. Maggiora, and K. C. Chou, *Biopolymers*, **31**, 1077 (1991).
34. M. Watanabe and M. Karplus, *J. Phys. Chem.*, **99**, 5680 (1995).
35. (a) M. Fixman, *Proc. Natl. Acad. Sci. USA*, **71**, 3050 (1974); (b) M. Fixman, *J. Chem. Phys.*, **69**, 1527 (1978); (c) W. F. Van Gunsteren, *Mol. Phys.*, **40**, 1015 (1980).